
# ***Autumn Examinations 2014/2105***

| | |
|---|---|
| **Exam Code(s)** | 4BCT |
| **Exam(s)** | 4th year Information Technology and Computer Science |
| **Module Code(s)** | CT422 |
| **Module(s)** | Modern Information Management |
| **Discipline(s)** | Information Technology |
| **Internal Examiner(s)** | Dr. Michael Madden<br>Dr. Colm O'Riordan* |
| **External Examiner(s)** | Dr. John Power |
| **No. of Pages** | 3 (including cover sheet) |
| **Duration** | 2 hours |
| **Instructions:** | Answer any three questions |

**PTO**

**Q.1**

i) The vector space model for Information Retrieval is one of the most commonly adopted models. Outline the model explaining both the representation of queries and documents and a means to calculate similarity. Discuss the advantages and disadvantages of such a model. *(9)*

ii) The accuracy of the vector space model depends on the quality of the weighting of the terms in both the query and documents. Discuss, with reference to well known weighting schemes, the main components of a good weighting scheme. *(8)*

iii) Describe suitable data structures you would use to implement an information retrieval system adopting the vector space model. Discuss the efficiency of your proposed approach. *(8)*

**Q.2**

i) Empirical evaluation of information retrieval systems plays an important role in information retrieval research. Define and discuss the following metrics that can be used to measure the performance of an Information Retrieval system: *precision*, *recall*, *novelty* and *coverage*. *(10)*

ii) The concepts of *topical relevance* and *component coverage* have been used to evaluate approaches to structured retrieval (e.g., retrieval of XML documents). Describe, with a suitable example, these concepts. *(7)*

iii) With reference to a clustering algorithm of your choice, describe suitable approaches to measuring the quality of the clustering algorithm. Your answer should distinguish between internal and external criteria. *(8)*

**Q.3**

i) Query modification is often used by systems to attempt to improve precision and recall for a given information need. Discuss an approach, given user feedback on the returned answer, to improve the performance of the query. *(8)*

ii) Query augmentation can also take place without explicit user feedback. Outline an approach to automatically generate suggested keywords for a user to augment their query. *(8)*

iii) Traditionally, information retrieval systems and web search engines have presented results in a list ordered by estimated relevance. Identify any limitations with this approach and suggest an alternative presentation of results. *(9)*

**Q.4**

i) Define what is meant by *collaborative filtering*.

Given the following matrix of ratings by four people for six items:

| | *a* | *b* | *c* | *d* | *e* | *f* |
|---|---|---|---|---|---|---|
| Rick | 2 | 5 | | 2 | 4 | |
| Rob | 4 | 2 | | 5 | 1 | 2 |
| Garth | 2 | 4 | 3 | | | 5 |
| Levon | 2 | 4 | | 5 | 1 | |

(Note: Ratings are in the range 1-5, where 1 indicates “dislike” and 5 indicates “like”).

Describe a neighbourhood based approach using Pearson Correlation and weighted averages to generate a prediction for Rick for item *f*. Your answer should:

- Describe the data given.
- Outline the main steps in the neighbourhood based approach.
- Show the steps required to predict a rating.

*(12)*

ii) Many modern web-based search engines attempt to take into account the web link structure in addition to the content of the pages. Outline an approach that could be adopted to use information embedded in the web link structure to return relevant documents to a user. Discuss any limitations associated with this approach. *(13)*